



TITLE:

# A Study on Web Search based on Coordinate Relationships( Abstract\_要旨)

AUTHOR(S):

Meng, Zhao

---

CITATION:

Meng, Zhao. A Study on Web Search based on Coordinate Relationships. 京都大学, 2016, 博士(情報学)

ISSUE DATE:

2016-09-23

URL:

<https://doi.org/10.14989/doctor.k20030>

RIGHT:

許諾条件により本文は2017-01-01に公開

( 続紙 1 )

京都大学	博士 ( 情報学 )	氏名	趙 夢
論文題目	A Study on Web Search based on Coordinate Relationships (同位関係に基づくウェブ検索に関する研究)		
(論文内容の要旨)			
<p>本論文は、従来の、情報内容の類似に基づく情報検索 (Content-Based Information Retrieval, CBIR) の弱点を補うための検索方式を、情報の同位関係という概念に基づき、開発したものである。CBIRは、テキスト・文書・画像などが検索クエリとして与えられ、それと内容的に類似性の高いテキスト・文書・画像などを検索するものである。CBIRは、画像検索においては社会的ニーズも高く、すでに一部の商用検索エンジンで実現されている。</p> <p>本論文では、従来のCBIRでは実現が困難であった、内容の「部分的な」類似性に基づく情報検索、換言すると「似て非なる」情報の検索行う手法を開発している。「似て非なる」情報検索をモデル化するために、本論文では、情報の同位関係 (coordinate relationship) に基づくモデルを提案している。従来、2つの語 (概念) に対して、両者が共に所属する上位概念が存在するとき、この2つの語 (概念) は同位関係にあると定義される。本論文では、この同位関係を、さらに、語集合、語の組にまで拡張している。語の組の間の同位関係をもとにして、文のパラフレーズ発見によるクエリ変換手法を開発している。さらに、語集合間の同位関係をもとにして、トピックが類似しているが事象が異なるような文書の検索を行う手法を開発している。さらに、画像内容の類似性と隣接性に着目した周辺類似画像の検索や、入力画像と景観類似する他地点の画像の検索を行う手法も開発している。</p> <p>本論文は7章からなる。その概要は以下の通りである。</p> <p>第1章では、まず、本研究の背景および動機について述べたあと、本研究の概要について述べている。</p> <p>第2章では、本研究に関係するこれまでの関連研究の概要について述べている。</p> <p>第3章では、本研究における中心となる概念である、同位関係 (coordinate relationship) について定義・説明を行っている。</p> <p>第4章は、文章そのものを検索クエリとする文章クエリ (sentential queries) のクエリ変換方式について研究を行ったもので、従来に比べて、検索性能、特に再現率を向上させる処理方式を提案している。提案方式は、与えられた文章クエリと意味的に同義の文章 (パラフレーズ) をWeb検索で求め、これらの文章で元の検索クエリを置き換えるものである。提案方式においては、文を関係表現であるとみなし、文テンプレートと値のnヶ組に分解し、文テンプレート間のパラフレーズ関係と、値のnヶ組間の同位関係を相互的に発見するための相互強化学習アルゴリズムを提案している。</p> <p>第5章は、文書間の同位関係にもとづき、与えられた文書と「似て非なる」関係にある文書を検索する手法を提案している。与えられた2文書が「似て非なる」関係にあるとは、内容的には類似性が高いが、それぞれが異なる事象や概念の記述であることを意味する。さらに、本章では、文書間の同位関係に基づいて、情報検索結果を構造化するための制約付きクラスタリング手法を提案している。</p> <p>第6章では、景観画像検索に関する検索手法を提案している。ある場所の写真画像を入力とし、その画像が有する「景観」 (Landscape) と類似する他の場所の画像を検索出力するものである。通常、都市景観は1枚の写真で表現することが難しいという点に注目し、入力画像の場所に隣接する場所も類似の景観を有する可能性が高いことに着目し、画像の類似性と画像の隣接性の双方を考慮に入れたPanoramaRankと呼ばれる画像検索手法を提案している。</p> <p>第7章は結論であり、本研究で得られた成果全般をまとめるとともに、今後の課題について述べている。</p>			

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、内容の類似性に基づく情報検索を行うための検索方式を、情報の同位関係という概念に基づき、提案したものである。内容の類似性に基づく情報検索は、テキスト、文書、画像自身などが検索クエリとして与えられ、検索クエリと内容的に類似性の高いテキスト、文書、画像などを検索するものである。従来の、内容の類似性に基づく情報検索は、与えられたクエリ内容の「全体的な」類似性に注目した情報検索であるが、ウェブ情報検索が普及した現在、クエリ内容の「部分的な」類似性に基づく情報検索のニーズが高まっている。例えば、具体的なWebページ（フランスワインについて記述したページ）を閲覧している際、類似した体裁でドイツワインについて記述したページを探したいという検索要求が起こることが多い。このような検索は、内容に基づく「似て非なる」(similar-but-different)情報検索と位置付けることができ、従来の検索手法では検索性能が低く、実現が困難なものであった。

本論文では、内容の「部分的な」類似性に基づく情報検索（「似て非なる」情報の検索）を行う手法を開発している。この「似て非なる」情報の検索をモデル化するために、本論文では、情報の同位関係 (coordinate relationship) に基づくモデルを提案している。従来、2つの語（概念）に対し、両者が共に所属する上位概念が存在するとき、この2つの語（概念）は同位関係にあると定義される。本論文では、この同位関係を、さらに、語集合、語の組にまで拡張している。

本論文で得られている研究成果の概要は以下の通りである。

1. 文章クエリ (sentential queries) によるWeb検索は、多くの検索エンジンが文章のパラフレーズ処理を行っていないため、これまで、高い再現率を得ることが困難であった。本論文では、文章クエリのパラフレーズ表現を検索時に求め、得られたパラフレーズ表現を新たな文章クエリとする、クエリ変換方式を開発している。提案方式は、文を文テンプレートと値の $n$ 組に分解し、文テンプレート間のパラフレーズ関係と、値の $n$ 組間の同位関係を、相互的に発見する相互強化学習アルゴリズムに基づくものである。さらに、提案方式の評価実験を行い、多くのパラフレーズ表現をWebから発見することに成功し、それによって、再現率を大幅に向上できることを検証している。
2. 文書間の同位関係にもとづき、与えられた文書から、トピックの類似性は高いが、異なる事象や概念の記述である文書を検索する手法、および、検索結果の文書集合を、文書間同位関係を保持しつつクラスタリングする手法を提案した。ニューヨークタイムズの記事コーパスを用いた評価実験により提案した検索方式の有効性を検証している。さらに、文書間の同位関係に基づき情報検索結果を構造化するための制約付きクラスタリング手法を提案した。この制約付きクラスタリング手法の有効性も、検索されたニュース記事集合に適用して評価実験を行い、良好なクラスタリングが行えることを確認している。
3. 景観画像検索に関する新規性・有効性の高い検索手法を提案している。ある場所の写真画像を入力とし、その画像が有する「景観」と類似する他の場所の画像を検索出力するもので、画像の類似性と画像の隣接性の双方を考慮に入れたPanoramaRankと呼ばれる画像検索手法を提案している。本手法は、本来、入力画像と「景観」が類似する隣接場所の画像を検索するものであるが、この手法を用いて、最終的に、「景観」が類似する、他の場所の画像を検索することが可能となる。提案手法の評価実験を行い、本手法の有効性を示している。

以上のとおり、本論文では、従来実現が困難であった、内容の「部分的な」類似性に基づく情報検索を行う手法を開発している。特に、語の組の間の同位関係に基

づく文のパラフレーズ発見、語集合間の同位関係に基づくトピック類似・異事象記述文書の検索、画像内容の類似性と隣接性に着目した周辺類似画像検索などに関し、新規性・有効性の高い手法を開発している。

このように、本論文は、学術上寄与するところが少なくないため、博士（情報学）の学位論文として価値あるものと認める。また、平成28年8月10日に論文内容とそれに関連した事項について試問を行った結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日： 年 月 日以降